UNIVERSITY OF
CAMBRIDGE

Cambridge Working Papers in Economics

EGARCH models with fat tails, skewness and leverage

Andrew Harvey and Genaro Sucarrat

CWPE 1236

EGARCH models with fat tails, skewness and leverage

Andrew Harvey and Genaro Sucarrat
Faculty of Economics, Cambridge University
and Department of Economics, BI Norwegian Business School, Oslo
ACH34@ECON.CAM.AC.UK

August 15, 2012

Abstract

An EGARCH model in which the conditional distribution is heavy-tailed and skewed is proposed. The properties of the model, including unconditional moments, autocorrelations and the asymptotic distribution of the maximum likelihood estimator, are obtained. Evidence for skewness in conditional t-distribution is found for a range of returns series and the model is shown to give a better fit than the corresponding skewed-t GARCH model.

Abstract

KEYWORDS: General error distribution; heteroskedasticity; leverage; score; Student's t, two components.

JEL classification; C22, G17

1 Introduction

An EGARCH model in which the variance, or scale, is driven by an equation that depends on the conditional score of the last observation was proposed by Creal, Koopman and Lucas (2008, 2011) and Harvey and Chakravarty

(2008).[1] The model has a number of attractions. In particular, an exponential link function ensures positive scale parameters and enables the conditions for stationarity to be obtained straightforwardly. Furthermore, although deriving a formula for the autocorrelation function (ACF) of squared observations is less straightforward than it is for a GARCH model, analytic expressions can be obtained and these expressions are more general. Specifically, formulae for the ACF of the absolute values of the observations raised to any power can be obtained. Finally, not only can expressions for multi-step forecasts of volatility be derived, but their conditional variances can be also found and the full conditional distribution is easily simulated.

When the conditional score is combined with an exponential link function, the asymptotic distribution of the maximum likelihood estimator of the dynamic parameters can be derived; see Harvey (2011). The theory is much more straightforward than it is for GARCH models. An analytic expression for the asymptotic covariance matrix can be obtained and the conditions for the asymptotic theory to be valid are easily checked.

A heavy-tailed conditional distribution can be modeled by a Student t-distribution, as in the GARCH-t model of Bollerslev (1987). However, the use of the conditional score in the dynamic volatility equation in what we call the Beta-t-EGARCH model means that observations that would be considered outliers for a Gaussian distribution are downweighted. An announcement made by the computer firm Apple illustrates the robustness of Beta-t-EGARCH. On Thursday 28 September 2000 a profit warning was issued,[2] which led the value of the stock to plunge from an end-of-trading value of \$26.75 to \$12.88 on the subsequent day. In terms of volatility this fall was a one-off event, since it apparently had no effect on the variability of the price changes on the following days. Figure 1, contains a snapshot of the event and the surrounding period. The figure plots absolute returns, the fitted conditional standard deviations of a GARCH(1,1)-t specification with leverage, and the fitted conditional standard deviations of the comparable Beta-t-EGARCH model; a full set of estimation results are given later in table 5. As is clear from the figure, the GARCH forecasts of one-step standard deviations exceed absolute returns for almost two months after the event, a clear-cut example of forecast failure. By contrast, the Beta-t-EGARCH

[1] Estimation and inference of the first-order Beta-t-EGARCH model is available via the R package betategarch, see Sucarrat (2012).

[2] CNN Money, see http://money.cnn.com/2000/09/29/markets/techwrap/. Retrieved 1 November 2011.

forecasts remain in the same range of variation as the absolute returns.

The main contribution of this paper is to extend conditional score models to skew distributions. Conditional skewness has important implications for asset pricing, as discussed in Harvey and Siddique (2000). Here, the emphasis is on the skew-t leading to a model that we call Beta-skew-t-EGARCH. However, the same approach works for the general error distribution and gives the Gamma-skew-GED-EGARCH model. The preferred specification is one in which skewness in the conditional distribution of y_t is combined with leverage in the specification of scale. A two-component model gives further gains in goodness of fit and is able to mimic the long memory pattern displayed in the autocorrelations of the absolute values.

The t-distribution is skewed using the method proposed by Fernandez and Steel (1998). The advantage of the FS approach compared with other skewing approaches is its computational and analytic tractability, conceptual simplicity and ease of application across a wide range of densities. The FS method has been adopted by a number of researchers, recent examples being Zhu and Zinde-Walsh (2009), Zhu and Galbraith (2010), and Gomez et al (2007). In the context of changing variance, Giot and Laurent(2003, 2004) show that a skewed-t GARCH model (with leverage) does very well in predicting Value-at-Risk (VaR). This model is available as an option in the G@RCH package of Laurent (2009).

The plan of the paper is as follows. Section 2 outlines the foundations of the Beta-t-EGARCH model, whereas section 3 introduces skewness. Section 4 introduces a martingale difference (MD) modification of the model of section 3, which ensures that the innovation is a MD. Section 5 briefly outlines how the Gamma-Skew-GED-EGARCH class of models is obtained along the same lines as the Beta-Skew-t-EGARCH class, when the conditional distribution is GED instead of t. Section 6 contains an extensive set of empirical applications, while section 7 briefly notes how a time-varying location can be accommodated in terms of a dynamic conditional score model. Section 8 concludes and outlines several possible extensions.

2 Beta-t-EGARCH

The Beta-t-EGARCH model is

$$y_t = \mu + \varepsilon_t \exp(\lambda_{t|t-1}), \quad t = 1, ..., T, \tag{1}$$

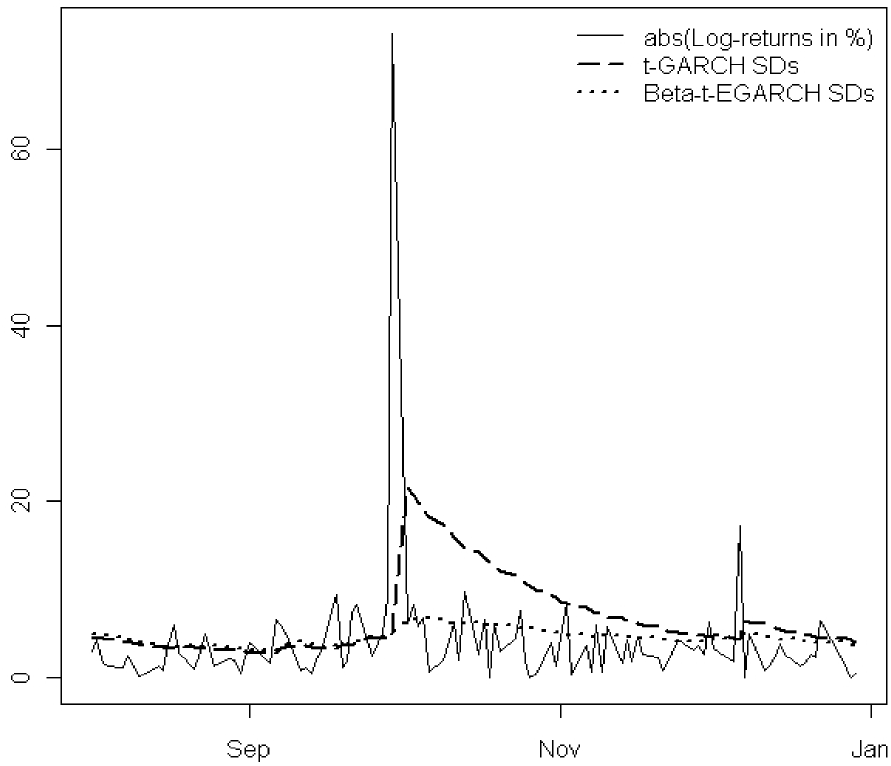

Figure 1: Apple returns with Beta-t-EGARCH and GARCH filters, both with leverage

where ε_t is a serially independent variable that has a $t_\nu-$distribution with positive degrees of freedom, ν, and $\lambda_{t|t-1}$, the logarithm of the scale, is a linear combination of past values of the conditional score

$$u_t = \frac{(\nu+1)(y_t-\mu)^2}{\nu\exp(2\lambda_{t|t-1})+(y_t-\mu)^2} - 1, \quad -1 \leq u_t \leq \nu, \quad \nu > 0. \tag{2}$$

The first-order model,

$$\lambda_{t+1|t} = \delta + \phi\lambda_{t|t-1} + \kappa u_t, \tag{3}$$

is stationary if $|\phi| < 1$. Since u_t is a martingale difference (MD) and hence WN, $\lambda_{t|t-1}$ is weakly stationary with an unconditional mean of $\omega = \delta/(1-\phi)$ and an unconditional variance of $\sigma_u^2/(1-\phi^2)$. Note that the process is assumed to have started in the infinite past, though for practical purposes $\lambda_{1|0}$ may be set equal to the unconditional mean. Identifiability requires $\kappa \neq 0$. Such a condition is hardly surprising since if κ were zero there would be no dynamics.

2.1 Moments and predictions

The conditional score may be expressed as

$$u_t = (\nu+1)b_t - 1, \quad t = 1, ..., T, \tag{4}$$

where, for finite degrees of freedom,

$$b_t = \frac{(y_t-\mu)^2/\nu\exp(2\lambda_{t|t-1})}{1+(y_t-\mu)^2/\nu\exp(2\lambda_{t|t-1})}, \qquad 0 \leq b_t \leq 1, \quad 0 < \nu < \infty, \tag{5}$$

is distributed as $beta(1/2, \nu/2)$ at the true parameter values. Since u_t depends on the same beta distribution in all time periods, it is independently and identically distributed (IID), not just a MD. It has zero mean and variance, $\sigma_u^2 = 2\nu/(\nu+3)$.

Harvey and Chakravarty (2008) derive expressions for the moments and autocorrelations of the observations. The odd moments of y_t are zero when the distribution of ε_t is symmetric. The even moments of y_t in the stationary

Beta-t-EGARCH model are

$$\begin{aligned} E[(y_t-\mu)^m] &= E(\varepsilon_t^m)E(\exp(m\lambda_{t|t-1})), \qquad (6) \\ &= \frac{\nu^{m/2}\Gamma(\frac{m}{2}+\frac{1}{2})\Gamma(\frac{-m}{2}+\frac{\nu}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})}e^{m\omega}\prod_{j=1}^{\infty}e^{-\psi_j m}\beta_\nu(\psi_j m), \quad m<\nu, \end{aligned}$$

where ψ_j, $j=1,2,..$ are the coefficients in the moving average representation,

$$\lambda_{t|t-1}=\omega+\sum_{j=1}^{\infty}\psi_j u_{t-j},$$

and $\beta_\nu(a)$ is Kummer's (confluent hypergeometric) function, ${}_1F_1(1/2;(\nu+1)/2;a(\nu+1))$; see Slater (1965, p 504).

Expressions for the autocorrelations of $\left|y_t-\mu_y\right|^c$, $c>0$, were also obtained. Note that

$$E(\exp(c\lambda_{t|t-1})=e^{c\omega}\prod_{j=1}^{\infty}e^{-\psi_j c}\beta_\nu(\psi_j c) \qquad (7)$$

is valid for any $c>0$.

The optimal predictor of scale in Beta-t-EGARCH is

$$E_T\left(e^{\lambda_{T+\ell|T+\ell-1}}\right)=e^{\lambda_{T+\ell|T}}\prod_{j=1}^{\ell-1}e^{-\psi_j}\beta_\nu(\psi_j), \qquad \nu>0, \quad \ell=2,3,.., \qquad (8)$$

where $\lambda_{T+\ell|T}$ is the linear predictor of $\lambda_{T+\ell|T+\ell-1}$. The MSE of the predicted scale for $\ell=2,3,...$, is

$$MSE(E_T\left(e^{\lambda_{T+\ell|T+\ell-1}}\right))=e^{2\lambda_{T+\ell|T}}\left(\prod_{j=1}^{\ell-1}e^{-2\psi_j}\beta_\nu(2\psi_j)-\left(\prod_{j=1}^{\ell-1}e^{-\psi_j}\beta_\nu(\psi_j)\right)^2\right).$$

The multi-step predictor of the variance of $y_{T+\ell}$ is obtained from the formula above with $Var(\varepsilon_t)$ included, that is

$$Var_T(y_{T+\ell})=\frac{\nu}{\nu-2}\left(\gamma^2-1+\gamma^{-2}\right)e^{2\lambda_{T+\ell|T}}\prod_{j=1}^{\ell-1}e^{-2\psi_j}\beta_\nu(2\psi_j), \quad \nu>2. \quad (9)$$

2.2 Asymptotic distribution of maximum likelihood estimator

The ML estimates are obtained by maximizing the log-likelihood function with respect to the unknown parameters. Although (3) is the conventional formulation of a first-order dynamic model, the information matrix takes a simpler form if the paramerization is in terms of ω rather than δ. Thus

$$\lambda_{t|t-1} = \omega + \lambda^{\dagger}_{t|t-1}, \quad \lambda^{\dagger}_{t+1|t} = \phi\lambda^{\dagger}_{t|t-1} + \kappa u_t, \qquad t = 1, ..., T. \tag{10}$$

Re-writing the above model in a similar way to (3) gives

$$\lambda_{t+1|t} = \omega(1-\phi) + \phi\lambda_{t|t-1} + \kappa u_t, \quad |\phi| < 1, \qquad t = 1, ..., T. \tag{11}$$

When ν and μ are known, the information matrix for a single observation is time-invariant and given by

$$\mathbf{I}(\boldsymbol{\psi}) = \sigma_u^2 \mathbf{D}(\boldsymbol{\psi})$$

where

$$\mathbf{D}(\boldsymbol{\psi}) = \mathbf{D}\begin{pmatrix} \widetilde{\kappa} \\ \widetilde{\phi} \\ \widetilde{\omega} \end{pmatrix} = \frac{1}{1-b}\begin{bmatrix} A & D & E \\ D & B & F \\ E & F & C \end{bmatrix} \tag{12}$$

with

$$A = \sigma_u^2, \qquad B = \frac{\kappa^2\sigma_u^2(1+a\phi)}{(1-\phi^2)(1-a\phi)}, \qquad C = \frac{(1-\phi)^2(1+a)}{1-a},$$

$$D = \frac{a\kappa\sigma_u^2}{1-a\phi}, \qquad E = c(1-\phi)/(1-a) \qquad and \qquad F = \frac{ac\kappa(1-\phi)}{(1-a)(1-a\phi)},$$

with

$$\begin{aligned} a &= \phi - \kappa\frac{2\nu}{\nu+3} \\ b &= \phi^2 - \phi\kappa\frac{4\nu}{\nu+3} + \kappa^2\frac{12\nu(\nu+1)(\nu+2)}{(\nu+7)(\nu+5)(\nu+3)} \\ c &= \kappa\frac{4\nu(1-\nu)}{(\nu+5)(\nu+3)}, \qquad \nu > 0. \end{aligned} \tag{13}$$

Recall that $\sigma_u^2 = 2\nu/(\nu+3)$. The key conditions for the limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}}-\boldsymbol{\psi})$ to be multivariate normal with zero mean vector and covariance matrix $\mathbf{I}^{-1}(\boldsymbol{\psi})$ are $\kappa \neq 0$ and $b < 1$. The proof is sketched out in the appendix.

The asymptotic distribution of $\widetilde{\boldsymbol{\psi}}$ is not affected when μ is estimated. Estimating ν does give a slight change since

$$Var(\boldsymbol{\psi},\nu) = \begin{bmatrix} \frac{2\nu}{\nu+3}\mathbf{D}(\boldsymbol{\psi}) & \frac{1}{(\nu+3)(\nu+1)}\begin{pmatrix} 0 \\ 0 \\ \frac{1-\phi}{1-a} \end{pmatrix} \\ \frac{1}{(\nu+3)(\nu+1)}\begin{pmatrix} 0 & 0 & \frac{1-\phi}{1-a} \end{pmatrix} & h(\nu)/2 \end{bmatrix}^{-1}, \qquad (14)$$

where $\mathbf{D}(\boldsymbol{\psi})$ is the matrix in (12) and

$$h(\nu) = \frac{1}{2}\psi'(\nu/2) - \frac{1}{2}\psi'((\nu+1)/2) - \frac{\nu+5}{\nu(\nu+3)(\nu+1)}, \qquad (15)$$

with $\psi'(.)$ being the trigamma function; see, for example, Taylor and Verblya (2004).

2.3 Monte Carlo experiments

Table 1 reports Monte Carlo results for the Beta-t-EGARCH model, (1) and (10) with μ known to be zero, but κ, ϕ, ω and ν unknown. The expression for the information matrix indicates that the asymptotic distribution of these parameters does not depend on the value of ω and this is supported by simulation evidence. For each experiment, which consisted of $N = 1000$ replications, the tables show the asymptotic standard error (ase) for each parameter, together with the numerical root mean square error (rmse).

For $T = 1000$, the ase underestimates the rmse. For κ the underestimation is rather small, at most 10%. For ω the bias seems to be in the other direction for ϕ close to one. Again the difference is rarely more than 10%. For ϕ the ase can be half the rmse when ϕ is 0.95 or 0.99, though the underestimation is less serious when κ is bigger.

The ase for ν is not very sensitive to the other parameters and the ratio of the ase to the rmse is around 0.65.

For $T = 10,000$, the ase's and rmse's for ω, ϕ and κ are all very close. For ν the ratio of the ase to the rmse is around to 0.8, so convergence to the asymptotic distribution is much slower. However, a zero mean static model

Table 1: Finite sample properties and the asymptotic standard errors of the Beta-t-EGARCH model: $y_t = \exp(\lambda_{t|t-1})\varepsilon_t, \quad \varepsilon_t \sim t_{\nu=6}, \quad \lambda_{t|t-1} = \omega + \lambda^{\dagger}_{t|t-1}, \quad \lambda^{\dagger}_{t|t-1} = \phi_1 \lambda^{\dagger}_{t-1|t-2} + \kappa_1 u_{t-1}$

Sample size T =1000:								
DGP (ω,ϕ_1,κ_1)	$rmse$ $(\hat{\omega})$	ase $(\hat{\omega})$	$rmse$ $(\hat{\phi})$	ase $(\hat{\phi})$	$rmse$ $(\hat{\kappa})$	ase $(\hat{\kappa})$	$rmse$ $(\hat{\nu})$	ase $(\hat{\nu})$
(0, 0.90,0.05)	0.053	0.049	0.075	0.052	0.016	0.016	1.357	0.844
(0, 0.90,0.10)	0.065	0.069	0.038	0.032	0.018	0.017	1.406	0.845
(0, 0.95,0.05)	0.069	0.069	0.058	0.024	0.014	0.013	1.334	0.844
(0, 0.95,0.10)	0.098	0.109	0.019	0.017	0.016	0.015	1.332	0.846
(0, 0.99,0.05)	0.198	0.226	0.010	0.006	0.010	0.010	1.371	0.845
(0, 0.99,0.10)	0.312	0.428	0.008	0.005	0.013	0.013	1.356	0.846
Sample size $T = 10,000$:								
DGP (ω,ϕ_1,κ_1)	$rmse$ $(\hat{\omega})$	ase $(\hat{\omega})$	$rmse$ $(\hat{\phi})$	ase $(\hat{\phi})$	$rmse$ $(\hat{\kappa})$	ase $(\hat{\kappa})$	$rmse$ $(\hat{\nu})$	ase $(\hat{\nu})$
(0, 0.90,0.05)	0.017	0.015	0.017	0.016	0.005	0.005	0.354	0.267
(0, 0.90,0.10)	0.022	0.022	0.010	0.010	0.006	0.005	0.336	0.267
(0, 0.95,0.05)	0.021	0.022	0.008	0.008	0.004	0.004	0.345	0.267
(0, 0.95,0.10)	0.032	0.034	0.005	0.005	0.005	0.005	0.325	0.267
(0, 0.99,0.05)	0.065	0.071	0.002	0.002	0.003	0.003	0.343	0.267
(0, 0.99,0.10)	0.118	0.135	0.002	0.002	0.004	0.004	0.317	0.268

Simulations (N = 1000 replications) in R version 2.13.2. $rmse$, root mean square error of estimates. ase, asymptotic standard error (computed as $T^{-1/2} \cdot (i_{jj}^{-1})^{1/2}$, where T is the sample size and (i_{jj}^{-1}) is element jj of the inverse of the information matrix). Estimation via the nlminb() function with upper and lower bounds on the parameter space equal to $(\infty, 0.999999999, \infty, \infty)$ and $(-\infty, -0.999999999, -\infty, 2.1)$, respectively. Initial values used: (0.005, 0.96, 0.02, 10).

gives similar results, so the DCS model is not displaying anything unusual.

The empirical distributions of the estimates for $T = 10,000$ showed no substantial deviations from normality.

2.4 Leverage

Leverage effects may be introduced into the model using the sign of the observations. For the first-order model, (3),

$$\lambda_{t|t-1} = \delta + \phi\lambda_{t-1|t-2} + \kappa u_{t-1} + \kappa^* sgn(-(y_{t-1} - \mu))(u_{t-1} + 1). \tag{16}$$

Taking the sign of *minus* $y_t - \mu$ means that the parameter κ^* is normally non-negative for stock returns. Although the statistical validity of the model does not require it, the restriction $\kappa \geq \kappa^* \geq 0$ may be imposed in order to ensure that an increase in the absolute values of a standardized observation does not lead to a decrease in volatility.

The expressions for moments and ACFs can be adapted to deal with leverage, as can the asymptotic thoery.

2.5 Two components

Alizadeh, Brandt and Diebold (2002, p 1088) argue strongly for two component (or two factor) stochastic volatility dynamics, in both equity and foreign exchange. They model such components using a SV framework while Engle and Lee (1999) proposed a two component GARCH model. In both papers, volatility is modeled with a long-run and a short-run component, the main role of the short-run component being to pick up the temporary increase in volatility after a large shock. Such a model can display long memory behaviour; see Andersen et al (2006, p 806-7).

The two-component model is

$$\lambda_{t|t-1} = \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}$$

where

$$\begin{aligned} \lambda_{1,t+1|t} &= \phi_1 \lambda_{1,t|t-1} + \kappa_1 u_t \\ \lambda_{2,t+1|t} &= \phi_2 \lambda_{2,t|t-1} + \kappa_2 u_t \end{aligned}$$

The model is easier to handle than the two-component GARCH model; see the discussion on the non-negativity constraints in Engle and Lee (1999, p 480).

In the DCS model, as with the GARCH model, the long-term component, $\lambda_{1,t|t-1}$, will usually have ϕ_1 close to one, or even set equal to one. The short-term component, $\lambda_{2,t|t-1}$, will typically have a higher κ combined with the lower ϕ. The model is not identifiable if $\phi_2 = \phi_1$. Imposing the constraint $0 < \phi_2 < \phi_1 < 1$ ensures identifiability and stationarity.

Table 2: Numerical properties of ML estimation of Beta-t-EGARCH in the case of unit root: $T = 10000$, $\nu = 6$, 1000 replications. Only ω and κ estimated (ϕ and ν fixed to 1 and 6, respectively)

DGP (ω, ϕ, κ)	$m(\hat{\omega})$	$s(\hat{\omega})$	$m(\hat{\kappa})$	$s(\hat{\kappa})$	$c(\hat{\omega}, \hat{\kappa})$
(0, 1, 0.05)	0.014	0.309	0.050	0.0027	0.0001
(0, 1, 0.10)	0.011	0.435	0.100	0.0038	0.0000

Simulations in R. $m(\cdot)$, average of estimates. $s(\cdot)$ and $c(\cdot, \cdot)$, sample standard deviation and sample covariance of estimates (division by N, not by $N-1$, where N is the number of replications). Estimation via the nlminb() function with upper and lower bounds on the parameter space equal to (∞, ∞) and $(-\infty, -\infty)$, respectively. Initial values used: (0.005, 0.02).

2.6 Nonstationarity

The EGARCH model is nonstationary when $\phi = 1$ in the first-order model as written in (10). When $\omega = \lambda_{1|0}$ is fixed and known, the result in sub-section 2.2 indicates that the limiting distribution of $\sqrt{T}(\widetilde{\kappa} - \kappa)$ is normal with mean zero and variance $(1-b)/\sigma_u^4$ (Since ω is given, estimating ν does not affect the asymptotic distribution of $\widetilde{\kappa}$.) For small κ, $Var(\widetilde{\kappa}) \simeq 2\kappa/\sigma_u^2$. Thus for a $t_\nu-$distribution the approximate standard error of $\widetilde{\kappa}$ is $\sqrt{\kappa(\nu+3)/\nu T}$, provided that $\kappa > 0$.

When the parameter ω is estimated, it appears from the simulation evidence in table 2 that the asymptotic distribution of the ML estimator of κ is unchanged. The approximate asymptotic standard errors for $\kappa = 0.05$ and 0.10 are 0.00274 and 0.00387 respectively and these are almost exactly the same as the values in table 2.

If ϕ is estimated unrestrictedly, it will have a non-standard distribution[3]. The simulations reported in table 3, where ω, ϕ and κ are all unknown parameters, indicate that the distribution of $\widetilde{\kappa}$ is unchanged, which is to be expected since, unlike $\widetilde{\phi}$, $\widetilde{\kappa}$ is not superconsistent. (The parameter ω is not estimated consistently but this should not affect the asymptotic distribution

[3] A reasonable conjecture is that the limiting distribution of $T\widetilde{\phi}$ can be expressed in terms of functionals of Brownian motion, as is the case when a series is a random walk and observations are regressed on their lagged values.

Table 3: Numerical properties of ML estimation of Beta-t-EGARCH in the case of an estimated unit root: $T = 10000$, $\nu = 6$. Thus ϕ, ω and κ estimated (and ν fixed to 6)

DGP: (ω, ϕ, κ)	$m(\hat{\omega})$	$s(\hat{\omega})$	$m(\hat{\phi})$	$s(\hat{\phi})$	$m(\hat{\kappa})$	$s(\hat{\kappa})$	$c(\hat{\omega}, \hat{\phi})$	$c(\hat{\omega}, \hat{\kappa})$
(0,1,0.05)	0.012	0.313	1.00	0.00033	0.050	0.0027	0.00000	0.00005
(0,1,0.10)	0.020	0.435	1.00	0.00031	0.100	0.0038	0.00000	-0.00006

(ω, ϕ, κ)	$c(\hat{\phi}, \hat{\kappa})$	$\hat{i}_{11}$	$\hat{i}_{12}$	$\hat{i}_{13}$	$\hat{i}_{22}$	$\hat{i}_{23}$	$\hat{i}_{33}$
(0,1,0.05)	0.00000	13.41	-1.046	-0.00705	932.7	-0.0141	0.00102
(0,1,0.10)	0.00000	6.90	5.308	0.00219	1059.8	0.0073	0.00053

Simulations in R (1000 replications). $m(\cdot)$, average of estimates. $s(\cdot)$ and $c(\cdot,\cdot)$, sample standard deviation and sample covariance of estimates (division by N, not by $N-1$, where N is the number of replications). $\hat{i}_{11}$, $\hat{i}_{12}$ and $\hat{i}_{22}$, estimates of the elements of the information matrix. Extreme observations were excluded from the computations in the second (23 observations in total) run of simulations, that is, when κ was equal to 0.1. Estimation via the nlminb() function with upper and lower bounds on the parameter space equal to (∞, ∞, ∞) and $(-\infty, -\infty, -\infty)$, respectively. Initial values used: (0.005, 0.96, 0.02).

of $\widetilde{\phi}$ and $\widetilde{\kappa}$.)

3 Skew distributions

Skewness may be introduced into the Beta-t-EGARCH model using the method proposed by Fernandez and Steel (1998). The first sub-section describes the Fernandez and Steel method and the remaining sub-sections present the details for Beta-t-EGARCH. The same methods can be used for Gamma-GED-EGARCH, as described in section 5.

3.1 Method of Fernandez and Steel

The skewing method proposed by Fernandez and Steel (1998) uses a continuous probability density function, $f(z)$, that is unimodal and symmetric about zero to construct a skewed probability density function

$$f(\varepsilon_t|\gamma) = \frac{2}{\gamma+\gamma^{-1}}\left[f\left(\frac{\varepsilon_t}{\gamma}\right)I_{[0,\infty)}(\varepsilon_t) + f(\varepsilon_t\gamma)I_{(-\infty,0)}(\varepsilon_t)\right], \tag{17}$$

where $I(\varepsilon_t)$ is an indicator variable, taking the value one when $\varepsilon_t \geq 0$ and zero otherwise, and γ is a parameter in the range $0 < \gamma < \infty$. An equivalent but more compact formulation is

$$f(\varepsilon_t|\gamma) = \frac{2}{\gamma+\gamma^{-1}}f\left(\frac{\varepsilon_t}{\gamma^{sgn(\varepsilon_t)}}\right). \tag{18}$$

Symmetry is attained when $\gamma = 1$, whereas $\gamma < 1$ and $\gamma > 1$ produce left and right skewness respectively. In other words the left hand tail is heavier when $\gamma < 1$.

The uncentered moments of ε_t, given by Fernandez and Steel (1998), are

$$E(\varepsilon_t^c) = M_c\frac{\gamma^{c+1}+(-1)^c/\gamma^{c+1}}{\gamma+\gamma^{-1}} \tag{19}$$

where

$$M_c = 2\int_0^\infty z^c f(z)dz = E(|z|^c). \tag{20}$$

Note that $\sigma_z^2 = Var(z_t) = M_2$. Hence

$$E(\varepsilon_t) = \mu_\varepsilon = M_1(\gamma - 1/\gamma), \tag{21}$$

which is not zero unless $\gamma = 1$, and

$$Var(\varepsilon_t) = M_2\left(\gamma^2 - 1 + \gamma^{-2}\right) - M_1^2(\gamma - 1/\gamma)^2 \tag{22}$$

The standard measure of skewness is

$$\begin{aligned} E(\varepsilon_t - \mu_\varepsilon)^3 &= E(\varepsilon_t^3) - 3\mu_\varepsilon E(\varepsilon_t^2) + 2\mu_\varepsilon^3 \\ &= (\gamma - \gamma^{-1})[(M_3 + 2M_1^3 - 3M_1M_2)(\gamma^2+\gamma^{-2}) + 3M_1M_2 - 4M_1^3] \end{aligned}$$

divided by $(Var(\varepsilon_t))^{3/2}$; see Fernandez and Steel (1998, eq 6).

The introduction of a location parameter, μ, and λ, the logarithm of scale, so that

$$y_t = \mu + \varepsilon_t \exp(\lambda),$$

gives

$$f(y_t|\gamma) = \frac{2}{\gamma + \gamma^{-1}} \left[f\left(\frac{y_t - \mu}{\gamma \exp(\lambda)}\right) I_{[0,\infty)}(y - \mu) + f\left(\frac{(y_t - \mu)\gamma}{\exp(\lambda)}\right) I_{(-\infty,0)}(y_t - \mu) \right] \tag{23}$$

As regards moments of the observations,

$$\mu_y = E(y_t) = \mu + \mu_\varepsilon \exp(\lambda),$$

while $Var(y_t) = E(y_t - \mu_y)^2 = Var(\varepsilon_t)\exp(2\lambda)$.

The median and mean are both less than μ when $\gamma < 1$, the former because $\Pr(y_t \leq \mu) = 1/(1+\gamma^2) > 0.5$ and the latter because $(\gamma - 1/\gamma) < 0$ in (21).

3.2 Beta-skew-t-EGARCH

When the conditional distribution of a Beta-t-EGARCH model, (1), is skewed, the log-density is

$$\begin{aligned} \ln f_t &= \ln 2 - \ln(\gamma + \gamma^{-1}) + \ln \Gamma\left((\nu+1)/2\right) - \frac{1}{2}\ln \pi - \ln \Gamma\left(\nu/2\right) - \frac{1}{2}\ln \nu \\ & -\lambda_{t|t-1} - \frac{(\nu+1)}{2} \ln\left(1 + \frac{(y_t - \mu)^2}{\gamma^{2\mathrm{sgn}(y_t-\mu)}\nu e^{2\lambda_{t|t-1}}}\right). \end{aligned} \tag{24}$$

The score is

$$u_t = u_t^+ I_{[0,\infty)}(y_t - \mu) + u_t^- I_{(-\infty,0)}(y_t - \mu), \quad t = 1, ..., T, \tag{25}$$

where $u_t = u_t^+$ and $u_t = u_t^-$ are as in (2), but with b_t defined as

$$b_t^+ = \frac{(y_t - \mu)^2/\nu\gamma^2 \exp(2\lambda_{t|t-1})}{1 + (y_t - \mu)^2/\nu\gamma^2 \exp(2\lambda_{t|t-1})} \qquad or \quad b_t^- = \frac{(y_t - \mu)^2/\nu\gamma^{-2} \exp(2\lambda_{t|t-1})}{1 + (y_t - \mu)^2/\nu\gamma^{-2} \exp(2\lambda_{t|t-1})}$$

depending on whether $y_t - \mu$ is non-negative (b_t^+) or negative (b_t^-). However, the properties of u_t^+ and u_t^- do not depend on the sign of $y_t - \mu$ since in both cases they are a linear function of a variable with the same beta distribution. Hence, as before, u_t is IID with zero mean and variance is $2\nu/(\nu+3)$.

3.3 Asymptotic distribution of maximum likelihood estimator

When γ is known and there is no leverage, the information matrix is exactly as in the symmetric case because the distribution of the score and its first derivative depend on IID beta variates with the same distribution.

The asymptotic distribution of the ML estimators of the dynamic parameters is affected when γ is also estimated by ML. Zhu and Galbraith (2010) give an analytic expression for the information matrix, but with a different parameterization for the scale and the skewing parameter, which is $\alpha = 1/(1+\gamma^2)$. Thus α is in the range 0 to 1 and symmetry is $\alpha = 0.5$. The scale measure is

$$\sigma = (\gamma + 1/\gamma)\sigma'/2 = (\gamma + 1/\gamma)\exp(\lambda)\sqrt{\nu/4(\nu-2)},$$

where σ' is the standard deviation in the FS model; see Zhu and Galbraith (2010, eq 4). The same result can be found in Gomez et al (2007, proposition 2.3). Our formulae for the information matrix may be adapted quite easily by re-defining λ as $\ln\sigma$. The full information matrix for the dynamic model is then constructed as in sub-section 2.2. The asymptotic theory still holds when skewness is combined with leverage, but the information matrix becomes more complicated.

A set of Monte Carlo experiments were run on the Beta-skew-t-EGARCH specification. The asymptotic theory indicates that the limiting distributions of ω, ϕ and κ are changed by the estimation of γ but the simulations indicated that any such changes were small. The inclusion of leverage makes no difference to the foregoing conclusion. The tables are available on request.

3.4 Moments and predictions

When the scale changes over time and the $m-th$ unconditional moment of y_t around μ exists, it may be written as in (6), but with $E\left(\varepsilon_t^m\right)$ now given by (19). Thus

$$\mu_y = Ey_t = \mu + \mu_\varepsilon E\left(e^{\lambda_{t|t-1}}\right) = \mu + M_1(\gamma - 1/\gamma)E\left(e^{\lambda_{t|t-1}}\right) \qquad (26)$$

and

$$\begin{aligned} Var(y_t) &= E[(y_t - \mu_y)^2] = E[(\varepsilon_t e^{\lambda_{t|t-1}} - \mu_\varepsilon E(e^{\lambda_{t|t-1}}))^2] \quad (27) \\ &= E(\varepsilon_t^2) E(e^{2\lambda_{t|t-1}}) - \mu_\varepsilon^2 (E(e^{\lambda_{t|t-1}}))^2. \end{aligned}$$

The expected value of the absolute value of a t_ν-variate raised to a power m is

$$E(|z|^m) = \frac{\nu^{m/2}\Gamma(\frac{m}{2}+\frac{1}{2})\Gamma(\frac{-m}{2}+\frac{\nu}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu}{2})} \quad (28)$$

This expression may be used to evaluate M_c in (20). The unconditional expectations, $E(\exp m\lambda_{t|t-1})$ are given by (7), just as in the symmetric case, because u_t in (25) depends on the same beta distribution. Thus, from (26), the mean of the observations is

$$\mu_y = \mu + \frac{\nu^{1/2}\Gamma((\nu-1)/2)}{\Gamma(\nu/2)\sqrt{\pi}}(\gamma - 1/\gamma)E(\exp(\lambda_{t|t-1})), \quad \nu > 1. \quad (29)$$

For $\nu > 2$, the unconditional variance is obtained as

$$Var(y_t) = \frac{\nu}{\nu-2}\left(\gamma^2 - 1 + \gamma^{-2}\right)E(e^{2\lambda_{t|t-1}}) - \left[\frac{\nu^{1/2}\Gamma((\nu-1)/2)}{\Gamma(\nu/2)\sqrt{\pi}}(\gamma - 1/\gamma)\right]^2 (E(e^{\lambda_{t|t-1}}))^2.$$

When the conditional distribution is skewed, the volatility may increase the skewness in unconditional distributions, just as it increases the kurtosis. The calculations can be carried out by evaluating

$$E[(y_t - \mu_y)^3] = E(\varepsilon_t^3)E(e^{3\lambda_{t|t-1}}) - 3\mu_\varepsilon E(\varepsilon_t^2)E(e^{\lambda_{t|t-1}})E(e^{2\lambda_{t|t-1}}) + 2\mu_\varepsilon^3 (E(e^{\lambda_{t|t-1}}))^2.$$

The skewness measure is then

$$S(\nu,\gamma) = \frac{E[(y_t - \mu_y)^3]}{\left[E[(y_t - \mu_y)^2]\right]^{3/2}} \quad (30)$$

and this may be compared with $E(\varepsilon_t - \mu_\varepsilon)^3/(Var(\varepsilon_t))^{3/2}$.

The ACF of $(y_t - \mu_y)^2$ can be obtained in the same way as for the symmetric model.

The multi-step predictor of the variance of $y_{T+\ell}$ given in (9) needs to be modified to

$$Var_T(y_{T+\ell}) = \frac{\nu}{\nu-2}\left(\gamma^2 - 1 + \gamma^{-2}\right) e^{2\lambda_{T+\ell|T}} \prod_{j=1}^{\ell-1} e^{-2\psi_j}\beta_\nu(2\psi_j) - (\mu_y - \mu)^2,$$

for $\ell = 2, 3, ...$ and $\nu > 2$. The formula for $\mu_y - \mu$ is given by (29).

3.5 Leverage

When there is leverage, as in (16), the dynamic equation becomes

$$\lambda_{t+1|t} = \omega(1-\phi) + \phi\lambda_{t|t-1} + \kappa u_t + \kappa^* sgn(-y_t + \mu)(u_t + 1).$$

In contrast to the symmetric model, $\lambda_{t+1|t}$ is no longer driven by a MD since the expectation of the variable in the last term is

$$E[sgn(y_t - \mu)(u_t + 1)] = (1-\gamma^2)/(1+\gamma^2) \tag{31}$$

because $E(u_t + 1) = 1$. The moments are adapted accordingly.

4 Modeling returns with the martingale difference modification

There is a problem with using the formulation of the previous Section for modeling returns because the conditional expectation,

$$E_{t-1}y_t = \mu + \mu_\varepsilon \exp(\lambda_{t|t-1}),$$

is not constant. Therefore y_t cannot be a MD. The solution is to let μ be time-varying. The model is re-formulated as

$$\begin{aligned} y_t &= \mu^S_{t|t-1} + \varepsilon_t \exp(\lambda_{t|t-1}), \quad t = 1, ..., T \\ \mu^S_{t|t-1} &= \mu_y - \mu_\varepsilon \exp(\lambda_{t|t-1}), \end{aligned} \tag{32}$$

where μ_y is a constant parameter, which is both the conditional and the unconditional mean. The time-varying parameter $\mu^S_{t|t-1}$ replaces μ in the likelihood function, (24). The score is now

$$u_t = \frac{(\nu+1)((y_t - \mu_y + \mu_\varepsilon \exp(\lambda_{t|t-1}))(y_t - \mu_y)}{\nu\gamma^{2\mathrm{sgn}(y_t - \mu_y + \mu_\varepsilon \exp(\lambda_{t|t-1}))} \exp(2\lambda_{t|t-1}) + (y_t - \mu_y + \mu_\varepsilon \exp(\lambda_{t|t-1}))^2} - 1. \tag{33}$$

Giot and Laurent (2003) transform their Skew-t GARCH model to make it a MD. They also standardize to make the variance one, but in our skew-t model this is not necessary.

4.1 Moments, skewness and volatility

The model in (32) can also be expressed as

$$y_t = \mu_y + (\varepsilon_t - \mu_\varepsilon)\exp(\lambda_{t|t-1}). \tag{34}$$

Since

$$E_{t-1}[(y_t - \mu_y)^2] = E_{t-1}[(\varepsilon_t - \mu_\varepsilon)^2 \exp(2\lambda_{t|t-1})],$$

it follows from the LIE that the unconditional variance of y_t is now

$$Var(y_t) = E[(y_t - \mu_y)^2] = Var(\varepsilon_t)E\exp(2\lambda_{t|t-1}),$$

but the fact that (33) does not have the simple beta distribution of (25) makes analytic evaluation more difficult.

The skewness in the MD model is

$$S(\nu,\gamma) = \frac{E[(\varepsilon_t - \mu_\varepsilon)^3]E\exp(3\lambda_{t|t-1})}{\left[E[(\varepsilon_t - \mu_\varepsilon)^2]E(\exp(2\lambda_{t|t-1}))\right]^{3/2}}$$

and so the factor by which skewness changes because of changing volatility is just

$$S_\nu = \frac{E\exp(3\lambda_{t|t-1})}{[E(\exp(2\lambda_{t|t-1}))]^{3/2}}, \qquad \nu > 3. \tag{35}$$

It follows from Hölder's inequality[4] that S_ν is greater than, or equal to, one.

[4] $E|x|^r \leq [E|x|^s]^{r/s}$. Here $x = \exp(\lambda) \geq 0$, and r and s can be set to 2 and 3 respectively.

4.2 Leverage effects

When there is leverage, the dynamic equation becomes

$$\lambda_{t+1|t} = \delta + \phi\lambda_{t|t-1} + \kappa u_t + \kappa^* sgn(-y_t + \mu_y - \mu_\varepsilon \exp(\lambda_{t|t-1}))(u_t + 1). \quad (36)$$

There is also a case for letting the leverage depend on $sgn(-y_t + \mu_y)$ so that (36) becomes

$$\lambda_{t+1|t} = \delta + \phi\lambda_{t|t-1} + \kappa u_t - \kappa^* sgn(y_t - \mu_y)(u_t + 1).$$

The rationale is that leverage should depend on whether the return is above or below the mean.

Leverage in itself does not induce skewness in the multi-step and unconditional distributions of Beta-t-EGARCH models. However, as was noted in the previous sub-section, when the conditional distribution is skewed, the volatility may increase the skewness in the unconditional distribution. The question then arises as to whether leverage exacerbates this increase.

4.3 Asymptotic theory

The expectation of u_t is zero, as it should be, since it can be written

$$\begin{aligned} u_t &= \frac{(\nu+1)(y_t - \mu_y + \mu_\varepsilon \exp(\lambda_{t|t-1}))^2 - (\nu+1)\mu_\varepsilon \exp(\lambda_{t|t-1})(y_t - \mu_y + \mu_\varepsilon \exp(\lambda_{t|t-1}))}{\nu \exp(2\lambda_{t|t-1})\gamma^{2\mathrm{sgn}(y_t - \mu_y + \mu_\varepsilon \exp(\lambda_{t|t-1}))} + (y_t - \mu_y + \mu_\varepsilon \exp(\lambda_{t|t-1}))^2} - 1 \\ &= \frac{(\nu+1){\varepsilon_t}^2 - (\nu+1)\mu_\varepsilon \exp(\lambda_{t|t-1})\varepsilon_t}{\nu \exp(2\lambda_{t|t-1})\gamma^{2\mathrm{sgn}(\varepsilon_t)} + {\varepsilon_t}^2} - 1 \\ &= (\nu+1)b_t - 1 - (\nu+1)\mu_\varepsilon[(1-b_t)\varepsilon_t \exp(-\lambda_{t|t-1})\nu^{-1}\gamma^{-2} I_{[0,\infty)}(\varepsilon_t) \\ &\quad +(1-b_t)\varepsilon_t \exp(-\lambda_{t|t-1})\nu^{-1}\gamma^2 I_{(-\infty,0)}(\varepsilon_t)]. \end{aligned}$$

Therefore

$$\begin{aligned} E(u_t) &= E[(\nu+1)b_t - 1] - (\nu+1)\mu_\varepsilon E[(1-b_t)\,|\varepsilon_t| \exp(-\lambda_{t|t-1})\nu^{-1}\gamma^{-1}]\gamma^{-1}(\gamma^2/(1+\gamma^2)) \\ &\quad -E[(1-b_t)\,|\varepsilon_t| \exp(-\lambda_{t|t-1})\nu^{-1}\gamma]\gamma(1/(1+\gamma^2)), \end{aligned}$$

which is zero as the first expectation is zero and the second and third expectations cancel.

The distribution of u_t does not depend on λ and the same is true of the distribution of its derivatives. The conditions for the ML estimator to be consistent and asymptotically normal hold just as they do in the symmetric case.

4.4 Forecasts

The quantile function of a skew t distribution is given by expression (9) in Giot and Laurent (2003). If the $\tau-$quantile is denoted as $skst(\tau, \nu, \gamma)$, the $\tau-$quantile of the one-step ahead predictive distribution of y_t is $\mu + e^{\lambda_{T+1|T}} skst(\tau, \nu, \gamma)$. Formulae for VaR (the same as the quantile formula) and expected shortfall in a skew-t are given in Zhu and Galbraith (2010, p. 300). These formulae may be used in one-step ahead prediction.

Formulae generalizing the multi-step ahead predictions of the volatility and observations, (8) and (9) respectively, for the symmetric Beta-t-EGARCH model are difficult to obtain. (Note that volatility has implications for skewness of multi-step distributions, just as it does for the unconditional distribution.) However, the main interest is in quantiles and the multi–step conditional distributions can be computed by simulation, simply by generating beta variates and combining them with an observation generated from a skew-t.

5 Gamma-skew-GED-EGARCH

In the Gamma-GED-EGARCH model, $y_t = \mu + \varepsilon_t \exp(\lambda_{t|t-1})$ and ε_t has a general error distribution (GED) with positive shape (tail-thickness) parameter υ and scale $\lambda_{t|t-1}$. The log-density function of the $t-$th observation is

$$\ln f_t(\boldsymbol{\alpha}, \upsilon) = -\left(1 + \upsilon^{-1}\right) \ln 2 - \ln \Gamma(1 + \upsilon^{-1}) - \lambda - \frac{1}{2} \left|y_t - \mu\right|^{\upsilon} \exp(-\lambda \upsilon),$$

leading to a model in which $\lambda_{t|t-1}$ evolves as a linear function of the score,

$$u_t = (\upsilon/2)(\left|y_t - \mu\right|^{\upsilon} / \exp(\lambda_{t|t-1}\upsilon) - 1, \quad t = 1, ..., T. \qquad (37)$$

Hence $\sigma_u^2 = \upsilon$. When $\lambda_{t|t-1}$ is stationary, the properties of the Gamma-GED-EGARCH model and the asymptotic covariance matrix of the ML estimators can be obtained in much the same way as those of Beta-t-EGARCH. The name Gamma-GED-EGARCH is adopted because $u_t = (\upsilon/2)\varsigma_t - 1$, where $\varsigma_t = |y_t - \mu|^{\upsilon} / \exp(\lambda_{t|t-1}\upsilon)$ has a $gamma(1/2,\, 1/\upsilon)$ distribution.

The model extends to the skew case in much the same way as does Beta–t-EGARCH. The asymptotic theory for a static model is set out in Zhu and Zinde-Walsh (2009).

6 Applications

In this section we fit various Beta-t-EGARCH specifications to a range of demeaned financial return series. The fit of these models is compared to that of the standard GARCH(1,1) model with a leverage term of the form proposed by Glosten, Jagannathan and Runkle (1993), henceforth GJR. Apart from one series, Apple, which was already studied in the introduction, all the data are contained in the period 1 January 1999 - 12 October 2011, which corresponds to a maximum of 3275 observations. But for some of the series the available number of data points is substantially smaller. Yahoo Finance (http://yahoo.finance.com/) is the source of the stock market indices, the stock prices and the gold price, whereas the European Central Bank (http://www.ecb.int/) and the US Energy Information Agency (http://www.eia.gov/) are the sources of the exchange rate data and the oilprice data, respectively.

Table 4 contains descriptive statistics of the returns series, and confirms that they exhibit the usual properties of excess kurtosis compared with the normal, and ARCH as measured by serial correlation in the squared returns. All of the stock returns—apart from DAX—and the oil return series exhibit negative skewness, whereas gold and the exchange rate returns exhibit positive skewness. (Below we will see that the unconditional positive skewness in DAX returns is converted into a negative conditional skewness when controlling for ARCH, GARCH and leverage.) For the exchange rate returns the positive skewness is presumably due to the fact that the more liquid currencies appear in the denominator of each of the three exchange rates: An increase in the exchange rate (say, EUR/USD) implies a depreciation in the less liquid currency (Euro) relative to the more liquid currency (USD). Only two series do not pass the test of whether returns are a MD at traditional significance levels, namely SP500 and Statoil. For this reason these two return series are demeaned by fitting AR(1) specifications with a constant, whereas the rest of the returns are demeaned by a constant only.

Table 4: Descriptive statistics of return series (January 1999 - October 2011)

	m	s	$Kurt$	$Skew$	MDH $[p-val]$	$ARCH_{20}$ $[p-val]$
Apple:	0.072	3.104	53.846	-1.964	0.03 [0.86]	36.18 [0.01]
SP500:	-0.001	1.364	10.061	-0.156	7.64 [0.01]	4357.63 [0.00]
Ftse:	-0.002	1.310	8.459	-0.121	2.16 [0.14]	3581.03 [0.00]
DAX:	0.006	1.623	6.926	0.023	0.33 [0.56]	2994.33 [0.00]
Nikkei:	-0.015	1.587	9.437	-0.377	0.86 [0.35]	3464.52 [0.00]
Boeing:	0.029	2.124	7.869	-0.185	0.06 [0.80]	806.82 [0.00]
Sony:	-0.044	2.184	8.524	-0.239	0.43 [0.51]	568.21 [0.00]
McDonald's:	0.034	1.701	7.754	-0.084	0.40 [0.53]	485.24 [0.00]
Merck:	-0.010	1.988	26.914	-1.429	0.11 [0.74]	41.19 [0.00]
Statoil:	0.073	2.414	7.703	-0.496	5.36 [0.02]	3888.85 [0.00]
EUR/USD:	0.005	0.671	5.451	0.067	0.06 [0.81]	583.21 [0.00]
GBP/EUR:	0.006	0.516	6.653	0.398	2.37 [0.12]	2186.80 [0.00]
NOK/EUR:	-0.004	0.444	10.801	0.253	2.26 [0.13]	1093.29 [0.00]
Oil:	0.070	2.426	7.712	-0.274	0.34 [0.56]	543.48 [0.00]
Gold:	0.151	3.211	7.218	0.189	0.08 [0.78]	1481.72 [0.00]

Notes: m, sample mean. s, sample standard deviation. $Kurt$, sample kurtosis. $Skew$, sample skewness. MDH, Escanciano and Lobato (2009) test for the Martingale Difference Hypothesis. $ARCH_{20}$, Ljung and Box (1979) test for serial correlation in the squared return.

Demeaned returns, y_t, are modeled as in section 4. The two-component specification is

$$\begin{aligned} y_t &= \exp(\lambda_{t|t-1})(\varepsilon_t - \mu_\varepsilon), \qquad \lambda_{t|t-1} = \omega + \lambda^\dagger_{1,t|t-1} + \lambda^\dagger_{2,t|t-1}, \\ \lambda^\dagger_{1,t|t-1} &= \phi_1 \lambda^\dagger_{1,t|t-1} + \kappa_1 u_{t-1}, \quad |\phi_1| < 1, \quad \phi_1 \neq \phi_2, \\ \lambda^\dagger_{2,t|t-1} &= \phi_2 \lambda^\dagger_{2,t|t-1} + \kappa_2 u_{t-1} + \kappa^* sgn(-y_{t-1})(u_{t-1}+1), \end{aligned}$$

with u_t as in (33) with $\mu_y = 0$. Following Engle and Lee (1999, p. 487) and others, only the short-term component has a leverage effect. A little experimentation indicated that this was a reasonable assumption to make here.

When only one component is used in the Beta-skew-t-EGARCH model it is comparable with the skewed-t-GJR, namely

$$\begin{aligned} y_t &= \sigma_{t|t-1} \widetilde{\varepsilon}_{t|t-1}, \quad t = 1, ..., T, \\ \sigma^2_{t|t-1} &= \omega + \phi_1 \sigma^2_{t-1|t-2} + \kappa_1 y^2_{t-1} + \kappa^* I(y_{t-1} < 0) y^2_{t-1}, \end{aligned}$$

where $\widetilde{\varepsilon}_t$ is a skewed t distribution with zero mean and unit variance. The use of $sgn(-y_{t-1})$ rather than the indicator $I(y_{t-1} < 0)$ makes no difference to the fit. Note that the persistence parameter in the GJR model is $\phi_1 + \kappa_1 + \kappa^*/2$, not ϕ_1; see Taylor (2005, p 221).

Tables 5 to 8 contain estimation results of the different financial returns. The results of the Apple data were used in the introduction to illustrate a drawback with the GARCH framework. The maximized likelihood of the Beta-skew-t-EGARCH model with leverage is clearly larger than that of the skewed-t-GJR model. The use of two components[5] gives a further improvement (the SC information criterion is lower for the two component model). Despite the large outlier, there is little evidence of negative skewness in the fit; the estimates of γ are greater than one.

All the results suggest that returns are fat-tailed[6] and the presence of either leverage or skewness (or both) is a common feature across a range of

[5]For some series, for example SP500, the estimate of κ_2 is less than that of κ^*, indicating that the short run effect of a large positive return is to reduce volatility. There may be plausible explanations, but if not, the constraint $\kappa_2 = \kappa^*$ may be imposed. When this was done here, there was usually a statistically significant decrease in the likelihood. However, the model still fitted well and there are no important implications regarding the overall merits of using two components.

[6]The maximum estimated value of the degrees of freedom parameter is 17 (FTSE).

series. In fact, the only return series in which neither leverage nor skewness is significant (at 10%) in any of the three fitted models is the EUR/USD exchange rate. For gold, whose market characteristics are very particular (the price is fixed twice a day rather than continuously as in other financial markets), leverage is significant only in the two-component model, and skewness is not significant in any of the three. Another notable feature is that the unconditional positive skewness in DAX returns is converted into negative and significant conditional skewness, when controlling for ARCH, GARCH and volatility asymmetry. All in all, the results thus provide broad support in favour of the Beta-skew-t-EGARCH, since according to the SC information criterion the GJR beats the βtE specifications in only three instances (Statoil, a Norwegian petroleum company, NOK/EUR and gold). Comparing the one-component and two-component versions of the Beta-skew-t-EGARCH (excluding the Apple stock where a longer sample is used for estimation), the two-component performs better according to SC in only two instances (FTSE and DAX).

Both leverage and negative skewness is particularly pronounced among the stock market indices. The leverage estimate is always positive, which yields the usual interpretation of large negative returns being followed by higher volatility. Similarly, the skewness parameter estimate ranges from 0.86 to 0.91, which means the risk of a large negative (demeaned) return is higher than a large positive (demeaned) return. Interestingly, but maybe not surprisingly, most of the large stocks with relatively regular earnings payouts (Apple, Boeing, Sony, McDonald's, Merck, Statoil) do not exhibit as much leverage or negative skewness as the indices, and sometimes the skewness is positive. A striking exception is Statoil whose negative skewness is 0.87.

As noted above the most liquid currency pair (EUR/USD) exhibits neither leverage nor skewness. This is in line with what might be expected. However, medium liquid exchange rates like EUR/GBP exhibit some skewness but no leverage, whereas relatively minor exchange rates like NOK/EUR exhibit substantial skewness and leverage. A common interpretation of "leverage" in an exchange rate context is that a large depreciation (for whatever reason) can induce higher volatility. This means the leverage parameter can be negative, since the sign depends on which currency is in the numerator of the exchange rate. Specifically, if the currency of the smaller economy is in the numerator, then one would expect a negative sign: A positive return means a depreciation in the smaller currency, which subsequently leads to an increase in volatility, and vice versa. This accounts for the negative and statistically significant leverage estimate of NOK/EUR.

Table 5: βtE and GJR specifications fitted to various return series (September 1984 - October 2011)

		$\hat{\omega}$ [se]	$\hat{\phi}_1$ [se]	$\hat{\phi}_2$ [se]	$\hat{\kappa}_1$ [se]	$\hat{\kappa}_2$ [se]	$\hat{\kappa}^*$ [se]	$\hat{\nu}$ [se]	$\hat{\gamma}$ [se]	$LogL$ (SC)	$ARCH(\widehat{u})$ [p−val]	$ARCH(\widehat{\varepsilon})$ [p−val]
Apple: (T=6835)	GJR	0.198 [0.059]	0.911 [0.017]		0.054 [0.010]		0.029 [0.014]	5.07 [0.30]	1.032 [0.016]	−16395.2 (4.805)	-	5.71 [0.99]
	βtE	0.782 [0.047]	0.986 [0.005]		0.038 [0.006]			5.24 [0.31]		−16374.8 (4.797)	42.77 [0.00]	18.51 [0.42]
	βtE	0.788 [0.042]	0.982 [0.005]		0.040 [0.006]		0.010 [0.003]	5.24 [0.31]		−16366.4 (4.795)	37.64 [0.00]	15.84 [0.54]
	βtE	0.778 [0.042]	0.982 [0.005]		0.040 [0.005]		0.009 [0.003]	5.25 [0.31]	1.031 [0.016]	−16364.5 (4.796)	38.39 [0.00]	15.90 [0.53]
	βtE	0.793 [0.081]	0.997 [0.001]	0.830 [0.050]	0.015 [0.003]	0.045 [0.007]		5.40 [0.33]		−16353.3 (4.793)	17.70 [0.34]	12.18 [0.73]
	βtE	0.791 [0.087]	0.998 [0.001]	0.862 [0.038]	0.014 [0.003]	0.041 [0.007]	0.020 [0.004]	5.42 [0.33]		−16341.9 (4.791)	19.29 [0.20]	12.88 [0.61]
	βtE	0.788 [0.087]	0.998 [0.001]	0.859 [0.039]	0.014 [0.003]	0.041 [0.007]	0.019 [0.004]	5.43 [0.33]	1.031 [0.016]	−16340.0 (4.792)	19.34 [0.20]	12.74 [0.62]
SP500: (T=3214)	GJR	0.017 [0.003]	0.917 [0.008]		0.000 [0.003]		0.146 [0.016]	10.09 [1.69]	0.872 [0.020]	−4761.7 (2.978)	-	18.95 [0.33]
	βtE	0.065 [0.115]	0.991 [0.003]		0.044 [0.005]			10.66 [1.86]		−4832.2 (3.017)	50.47 [0.00]	36.78 [0.01]
	βtE	−0.115 [0.051]	0.988 [0.002]		0.021 [0.004]		0.036 [0.003]	11.32 [1.71]		−4762.3 (2.976)	56.77 [0.00]	28.84 [0.04]
	βtE	0.145 [0.075]	0.988 [0.002]		0.027 [0.004]		0.039 [0.003]	11.73 [1.83]	0.860 [0.020]	−4740.9 (2.965)	58.46 [0.00]	27.99 [0.05]
	βtE	0.099 [0.144]	0.996 [0.003]	0.968 [0.022]	0.025 [0.012]	0.020 [0.012]		10.83 [1.92]		−4831.3 (3.021)	51.49 [0.00]	37.78 [0.00]
	βtE	0.016 [0.121]	0.995 [0.003]	0.957 [0.013]	0.028 [0.011]	−0.011 [0.013]	0.047 [0.005]	10.59 [1.64]		−4753.2 (2.975)	58.27 [0.00]	34.08 [0.00]
	βtE	0.114 [0.121]	0.997 [0.002]	0.975 [0.008]	0.016 [0.007]	0.009 [0.007]	0.044 [0.004]	11.01 [1.71]	0.867 [0.021]	−4735.2 (2.967)	57.28 [0.00]	31.47 [0.01]
FTSE: (T=3227)	GJR	0.021 [0.004]	0.899 [0.010]		0.007 [0.009]		0.160 [0.020]	14.80 [3.09]	0.876 [0.023]	−4731.9 (2.948)	-	30.24 [0.02]
	βtE	0.127 [0.078]	0.986 [0.003]		0.034 [0.004]		0.041 [0.004]	14.60 [2.89]	0.853 [0.022]	−4714.9 (2.937)	38.25 [0.00]	25.02 [0.09]
	βtE	0.133 [0.134]	0.993 [0.004]	0.941 [0.014]	0.031 [0.009]	−0.001 [0.011]	0.054 [0.005]	17.24 [4.08]	0.866 [0.022]	−4703.2 (2.935)	35.08 [0.00]	30.58 [0.01]

βtE, Beta-skew-t-EGARCH specification. GJR, Glosten et al. (1993) specification with skew-t density. (se), standard error of parameter estimate. T, number of observations. LogL, log-likelihood. SC, Schwarz (1978) information criterion computed as SC $= -2\text{LogL}/T + k(\ln T)/T$ where k is the number of estimated parameters in the log-volatility specification. $ARCH(\widehat{u}_t)$ and $ARCH(\widehat{\varepsilon}_t)$, Ljung and Box (1979) test for 20th. order serial correlation of the $\widehat{u}_t$ and the squared standardised residuals $\widehat{\varepsilon}_t^2$, respectively. The variance-covariance matrix is computed as $(-\hat{H})^{-1}$, where $\hat{H}$ is the numerically estimated Hessian.

Table 6: βtE and GJR specifications fitted to various return series series (January 1999 - October 2011)

		$\hat{\omega}$ $[se]$	$\hat{\phi}_1$ $[se]$	$\hat{\phi}_2$ $[se]$	$\hat{\kappa}_1$ $[se]$	$\hat{\kappa}_2$ $[se]$	$\hat{\kappa}^*$ $[se]$	$\hat{\nu}$ $[se]$	$\hat{\gamma}$ $[se]$	$LogL$ (SC)	$ARCH(\widehat{u})$ $[p-val]$	$ARCH(\widehat{\varepsilon})$ $[p-val]$
DAX: (T=3256)	GJR	0.032 [0.006]	0.898 [0.010]		0.019 [0.008]		0.144 [0.019]	11.99 [2.14]	0.890 [0.022]	−5530.8 (3.412)	-	57.59 [0.00]
	βtE	0.364 [0.082]	0.984 [0.003]		0.041 [0.004]		0.036 [0.004]	13.99 [2.79]	0.871 [0.021]	−5519.5 (3.405)	51.57 [0.00]	61.59 [0.00]
	βtE	0.571 [0.406]	0.995 [0.007]	0.933 [0.014]	0.041 [0.010]	−0.008 [0.013]	0.051 [0.005]	14.56 [3.18]	0.890 [0.022]	−5504.1 (3.401)	47.43 [0.00]	38.75 [0.00]
Nikkei: (T=3135)	GJR	0.054 [0.011]	0.889 [0.013]		0.030 [0.010]		0.115 [0.020]	13.47 [2.84]	0.912 [0.023]	−5439.3 (3.485)	-	15.83 [0.54]
	βtE	0.266 [0.051]	0.972 [0.005]		0.043 [0.005]		0.029 [0.004]	12.72 [2.36]	0.910 [0.023]	−5432.4 (3.481)	31.38 [0.02]	20.12 [0.27]
	βtE	0.259 [0.089]	0.994 [0.004]	0.932 [0.018]	0.021 [0.006]	0.021 [0.008]	0.037 [0.005]	13.31 [2.59]	0.912 [0.023]	−5424.9 (3.481)	34.26 [0.00]	27.25 [0.03]
Boeing: (T=3216)	GJR	0.055 [0.015]	0.926 [0.012]		0.034 [0.010]		0.057 [0.016]	7.25 [0.83]	0.995 [0.025]	−6576.0 (4.105)	-	34.24 [0.01]
	βtE	0.538 [0.073]	0.988 [0.004]		0.032 [0.005]		0.017 [0.003]	7.52 [0.88]	0.983 [0.024]	−6568.7 (4.1)	25.20 [0.09]	45.18 [0.00]
	βtE	0.599 [0.151]	0.997 [0.002]	0.949 [0.021]	0.017 [0.005]	0.019 [0.008]	0.024 [0.005]	7.69 [0.91]	0.989 [0.024]	−6564.8 (4.103)	22.60 [0.09]	42.93 [0.00]
Sony: (T=2270)	GJR	0.062 [0.026]	0.944 [0.015]		0.029 [0.010]		0.030 [0.015]	5.78 [0.67]	1.064 [0.029]	−4742.8 (4.199)	-	16.55 [0.48]
	βtE	0.462 [0.075]	0.986 [0.006]		0.031 [0.007]		0.008 [0.004]	5.81 [0.67]	1.064 [0.028]	−4739.7 (4.196)	22.78 [0.16]	20.07 [0.27]
	βtE	0.467 [0.101]	0.995 [0.003]	0.884 [0.102]	0.018 [0.006]	0.026 [0.012]	0.010 [0.006]	5.92 [0.69]	1.068 [0.028]	−4737.9 (4.202)	18.13 [0.26]	15.23 [0.43]

Notes: See table 5.

Table 7: βtE and GJR specifications fitted to various return series series (January 1999 - October 2011)

		$\hat{\omega}$ [se]	$\hat{\phi}_1$ [se]	$\hat{\phi}_2$ [se]	$\hat{\kappa}_1$ [se]	$\hat{\kappa}_2$ [se]	$\hat{\kappa}^*$ [se]	$\hat{\nu}$ [se]	$\hat{\gamma}$ [se]	$LogL$ (SC)	$ARCH(\widehat{u})$ [$p-val$]	$ARCH(\widehat{\varepsilon})$ [$p-val$]
McDonald's: (T=3216)	GJR	0.020 [0.006]	0.943 [0.008]		0.032 [0.008]		0.040 [0.016]	6.13 [0.62]	1.001 [0.024]	−5828.1 (3.639)	-	21.45 [0.21]
	βtE	0.269 [0.091]	0.991 [0.003]		0.034 [0.005]		0.014 [0.004]	6.22 [0.61]	0.993 [0.024]	−5813.0 (3.63)	20.15 [0.27]	24.18 [0.11]
	βtE	0.303 [0.138]	0.995 [0.003]	0.825 [0.098]	0.029 [0.005]	0.013 [0.013]	0.030 [0.008]	6.31 [0.62]	1.003 [0.024]	−5809.0 (3.633)	16.24 [0.37]	23.62 [0.07]
Merck: (T=3216)	GJR	0.126 [0.044]	0.867 [0.032]		0.076 [0.022]		0.051 [0.027]	4.59 [0.35]	0.967 [0.022]	−6167.9 (3.851)	-	0.54 [1.00]
	βtE	0.326 [0.076]	0.987 [0.004]		0.039 [0.007]		0.024 [0.004]	4.66 [0.34]	0.949 [0.023]	−6116.2 (3.819)	17.04 [0.45]	0.46 [1.00]
	βtE	0.312 [0.106]	0.996 [0.002]	0.963 [0.021]	0.015 [0.006]	0.029 [0.010]	0.028 [0.005]	4.70 [0.34]	0.955 [0.023]	−6114.6 (3.823)	14.08 [0.52]	0.37 [1.00]
Statoil: (T=2521)	GJR	0.082 [0.024]	0.940 [0.011]		0.024 [0.011]		0.037 [0.016]	10.39 [1.90]	0.866 [0.026]	−5428.3 (4.325)	-	18.62 [0.35]
	βtE	0.717 [0.069]	0.988 [0.004]		0.024 [0.004]		0.014 [0.003]	10.92 [2.02]	0.864 [0.026]	−5429.9 (4.326)	24.71 [0.10]	20.55 [0.25]
	βtE	0.693 [0.092]	0.993 [0.003]	0.920 [0.034]	0.022 [0.006]	0.001 [0.009]	0.023 [0.005]	11.77 [2.32]	0.870 [0.026]	−5426.0 (4.33)	25.96 [0.04]	24.77 [0.05]
EUR/USD: (T=3274)	GJR	0.002 [0.000]	0.966 [0.004]		0.029 [0.005]		0.004 [0.004]	11.50 [2.00]	1.003 [0.024]	−3133.8 (1.929)	-	11.42 [0.83]
	βtE	−0.549 [0.082]	0.995 [0.002]		0.018 [0.003]		0.001 [0.002]	11.67 [2.02]	1.003 [0.024]	−3131.6 (1.928)	13.92 [0.67]	11.31 [0.84]
	βtE	−0.548 [0.082]	0.994 [0.003]	0.582 [0.333]	0.021 [0.004]	−0.021 [0.010]	−0.006 [0.008]	11.38 [1.93]	1.005 [0.024]	−3129.1 (1.931)	9.74 [0.84]	12.93 [0.61]

Notes: See table 5.

Table 8: βtE and GJR specifications fitted to various return series series (January 1999 - October 2011)

		$\hat{\omega}$ [se]	$\hat{\phi}_1$ [se]	$\hat{\phi}_2$ [se]	$\hat{\kappa}_1$ [se]	$\hat{\kappa}_2$ [se]	$\hat{\kappa}^*$ [se]	$\hat{\nu}$ [se]	$\hat{\gamma}$ [se]	$LogL$ (SC)	$ARCH(\hat{u})$ [$p-val$]	$ARCH(\hat{\varepsilon})$ [$p-val$]
GBP/EUR: (T=3274)	GJR	0.001 [0.000]	0.946 [0.007]		0.049 [0.007]		0.004 [0.006]	10.88 [1.75]	1.064 [0.026]	−2055.8 (1.271)	-	7.97 [0.97]
	βtE	−0.872 [0.105]	0.994 [0.002]		0.028 [0.004]		−0.003 [0.003]	11.32 [1.80]	1.060 [0.026]	−2052.0 (1.268)	11.85 [0.81]	9.27 [0.93]
	βtE	−0.870 [0.127]	0.997 [0.002]	0.966 [0.022]	0.016 [0.005]	0.014 [0.006]	−0.003 [0.003]	11.55 [1.88]	1.057 [0.026]	−2050.8 (1.273)	11.44 [0.72]	8.78 [0.89]
NOK/EUR: (T=3274)	GJR	0.004 [0.001]	0.920 [0.011]		0.062 [0.010]		0.000 [0.003]	7.01 [0.86]	1.118 [0.026]	−1552.7 (0.963)	-	29.75 [0.03]
	βtE	−1.030 [0.053]	0.977 [0.007]		0.040 [0.006]		−0.018 [0.004]	7.49 [0.94]	1.123 [0.026]	−1554.2 (0.964)	21.28 [0.21]	56.05 [0.00]
	βtE	−1.036 [0.072]	0.989 [0.005]	0.796 [0.088]	0.028 [0.006]	0.020 [0.010]	−0.037 [0.008]	7.61 [0.96]	1.114 [0.026]	−1546.8 (0.965)	20.85 [0.14]	53.83 [0.00]
Oil: (T=3240)	GJR	0.100 [0.032]	0.942 [0.012]		0.022 [0.009]		0.035 [0.014]	8.28 [1.10]	0.932 [0.023]	−7196.5 (4.457)	-	30.01 [0.03]
	βtE	0.759 [0.062]	0.989 [0.004]		0.021 [0.004]		0.014 [0.003]	8.84 [1.21]	0.918 [0.023]	−7193.4 (4.455)	24.08 [0.12]	40.26 [0.00]
	βtE	0.733 [0.073]	0.992 [0.004]	0.840 [0.043]	0.021 [0.004]	0.010 [0.008]	0.034 [0.007]	9.01 [1.25]	0.935 [0.022]	−7186.9 (4.456)	18.47 [0.24]	26.30 [0.03]
Gold: (T=2318)	GJR	0.042 [0.022]	0.959 [0.007]		0.034 [0.006]		0.005 [0.006]	7.42 [1.08]	1.038 [0.028]	−5733.5 (4.967)	-	23.05 [0.15]
	βtE	0.892 [0.109]	0.994 [0.003]		0.023 [0.004]		−0.003 [0.003]	8.03 [1.28]	1.032 [0.028]	−5736.7 (4.97)	22.68 [0.16]	32.85 [0.01]
	βtE	0.882 [0.111]	0.994 [0.003]	0.075 [0.351]	0.023 [0.004]	0.014 [0.015]	0.027 [0.011]	7.75 [1.20]	1.031 [0.028]	−5733.5 (4.974)	22.06 [0.11]	36.81 [0.00]

Notes: See table 5.

7 Changing location

Returns sometimes exhibit mild serial correlation. Such effects may be removed prior to fitting a volatility model as was done in the previous section. However, rather than simply using a standard procedure for estimating an ARMA model, a DCS model may be fitted, thereby providing protection against outliers. Indeed a DCS model with a skew distribution may be fitted and location and volatility estimated jointly.

Another possibility to consider is that the serial correlation may actually arise as a consequence of combining serial correlation in scale with conditional skewness.

7.1 Joint estimation of location and scale

When $y_t \mid Y_{t-1}$ has a symmetric t_ν-distribution and the location changes over time, but the scale is constant, it may be captured by a model in which $\mu_{t|t-1}$ is generated by a linear function of

$$u_t^\mu = \left(1 + \frac{(y_t - \mu_{t|t-1})^2}{\nu \exp(-2\lambda)}\right)^{-1} v_t, \quad t = 1, ..., T, \quad \nu > 0, \tag{38}$$

where $v_t = y_t - \mu_{t|t-1}$ is the prediction error. The role of the term in parentheses in (38) is to downweight extreme observations. The variable can be written

$$u_t^\mu = (1 - b_t)(y_t - \mu_{t|t-1}), \tag{39}$$

where

$$b_t = \frac{(y_t - \mu_{t|t-1})^2 / \nu \exp(2\lambda)}{1 + (y_t - \mu_{t|t-1})^2 / \nu \exp(2\lambda)}, \qquad 0 \le b_t \le 1, \quad 0 < \nu < \infty, \tag{40}$$

is distributed as $beta(1/2, \nu/2)$. Hence the mean of u_t^μ is zero, as it should be.

The first-order model is

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t = \mu_{t|t-1} + \exp(\lambda_{t|t-1})\varepsilon_t, \quad t = 1, ..., T \\ \mu_{t+1|t} &= \delta + \phi\mu_{t|t-1} + \kappa u_t^\mu, \end{aligned} \tag{41}$$

This model might be interpreted as an approximation to an AR(1) process plus t-distributed white noise. More generally, a linear dynamic model of order (p, r) may be defined as

$$\mu_{t+1|t} = \delta + \phi_1 \mu_{t|t-1} + ... + \phi_p \mu_{t-p+1|t-p} + \kappa_0 u_t^{\mu} + \kappa_1 u_{t-1}^{\mu} + ... + \kappa_r u_{t-r}^{\mu}, \quad (42)$$

where $p \geq 0$ and $r \geq 0$ are finite integers and $\delta, \phi_1, .., \phi_p, \kappa_0, .., \kappa_r$ are (fixed) parameters. Stationarity (both strict and covariance) of $\lambda_{t|t-1}$ requires that the roots of the autoregressive polynomial lie outside the unit circle, as in an autoregressive-moving average model.

When the conditional distribution is skew-t,

$$u_t^{\mu} = u_t^{+} I_{[0,\infty)}(y_t - \mu_{t|t-1}) + u_t^{-} I_{(-\infty,0)}(y_t - \mu_{t|t-1}), \quad t = 1, ..., T, \quad (43)$$

where $u_t = u_t^{+}$ and $u_t = u_t^{-}$ are as in (39), but with b_t defined as

$$b_t^{+} = \frac{(y_t - \mu_{t|t-1})^2/\nu \exp(2\lambda)}{1 + (y_t - \mu_{t|t-1})^2/\nu\gamma^2 \exp(2\lambda)} \quad or \quad b_t^{-} = \frac{(y_t - \mu_{t|t-1})^2/\nu \exp(2\lambda)}{1 + (y_t - \mu_{t|t-1})^2/\nu\gamma^{-2} \exp(2\lambda)}, \quad (44)$$

depending on whether $y_t - \mu_{t|t-1}$ is non-negative (b_t^{+}) or negative (b_t^{-}). The properties of u_t^{+} and u_t^{-} do not depend on the sign of $y_t - \mu_{t|t-1}$ since in both cases they are a linear function of the same beta variable, as defined in (40). The asymptotic distribution of the ML estimators may be obtained.

Location and scale may be estimated jointly. The dynamic equations have the same form as before. Thus u_t^{μ} is defined as in (43) but with λ replaced in (44) by $\lambda_{t|t-1}$. Similarly μ_y is replaced by $\lambda_{t|t-1}$ in the various formulae for u_t. Both u_t and u_t^{μ} are MDs, dependent on beta variables with the same distribution. However, the unconditional information matrix cannot be evaluated in the same way as before because the variance of the score with respect to the location depends on the scale.

The case for adopting the MD modification of section 4 may not be so strong when there is serial correlation in the level. If the modification is to be made, then

$$\mu_{t|t-1}^{S} = \mu_{t|t-1} - \mu_{\varepsilon} \exp(\lambda_{t|t-1}),$$

where $\lambda_{t|t-1}$ from (41) replaces the constant mean μ_y in (32). Of course if the serial correlation is first removed by pre-filtering the MD model is appropriate.

8 Conclusions and extensions

This article shows that much of the theory for the basic Beta-t-EGARCH model generalizes to a skew-t model. Thus expressions may be obtained for unconditional moments of the observations and for predictions. An analytic expression can be derived for the information matrix of a first-order model and its structure gives insight into the way in which the estimators of parameters interact for different parameterizations. For example, if the dynamic equation is set up in terms of the mean, the asymptotic distribution is independent of its value. The effect of the skewness parameter may be similarly explored. Having said that, the derivation of an analytic expression for the information matrix of the ML estimators for the preferred specification, which is one that retains the martingale difference property, is more difficult.

The fact that a comprehensive set of theoretical properties can be derived for Beta-t-EGARCH models is a considerable attraction. Even more important, from the practical point of view, is that our results provide yet more evidence on the better fit afforded by the Beta-t-EGARCH specification as compared with the GARCH-t benchmark; see also the results in Harvey and Chakravarty (2008) and Creal, Koopman and Lucas (2011). The skew-t model with two components, the short-term one with a leverage effect, seemed to give the best results overall. We find both leverage and negative skewness to be particularly pronounced among stock market indices, such as SP 500, FTSE, DAX and Nikkei.

Zhu and Galbraith (2010) consider an asymmetric skew t-distribution in which the degrees of freedom takes on a different value according to the sign of the deviation from the mean. The Beta-skew-t-EGARCH model could in principle be extended in this way. There is also the possibility of introducing skewness into the multivariate model of Creal, Koopman and Lucas (2011). Zhang et al (2011) propose such a multivariate model based on the generalized hyperbolic distribution, but, as they note, computing the information matrix for this distribution is analytically intractable so deriving asymptotic properties of ML estimators using the methods employed in this paper will not be possible.

REFERENCES

Alizadeh, S., Brandt, M. and Diebold, F.X. (2002). Range-Based Estimation of Stochastic Volatility Models. Journal of Finance 57, 1047-1092.

Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X.: (2006). Volatility and correlation forecasting. In: Elliot, G., Granger, C., Timmerman, A. (Eds.), Handbook of Economic Forecasting, 777-878. Amsterdam: North Holland.

Brandt, A. (1986). The stochastic equation Yn+1 = AnYn + Bn with stationary coefficients. *Advances in Applied Probability* **18**, 211–220.

Creal, D., Koopman, S.J., and A. Lucas: (2008). A general framework for observation driven time-varying parameter models, Tinbergen Institute Discussion Paper, TI 2008-108/4, Amsterdam.

Creal, D., Koopman, S.J., and A. Lucas (2012). Generalized autoregressive score models with applications. J of Applied Econometrics (to appear).

Creal, D., Koopman, S.J. and A. Lucas (2011). A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations, Journal of Business and Economic Statistics, 29, 552-63.

Escanciano, J. C. and I. N. Lobato (2009). An automatic portmanteau test for serial correlation. Journal of Econometrics 151, 140-149.

Engle, R.F., and Lee, G.G.J. (1999). A Long-Run and a Short-Run Component Model of Stock Return Volatility. In R.F. Engle and H. White (eds.) Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive Granger. Oxford: Oxford University Press.

Fernández, C., and Steel, M. (1998). On Bayesian Modelling of Fat Tails and Skewness. Journal of the American Statistical Association 93, 359-371.

Giot, P. and Laurent, S. (2003). Value at Risk for Long and Short Trading Positions. Journal of Applied Econometrics 18, pp. 641-664.

Giot, P, and Laurent, S. (2004). Modelling Daily Value-at-Risk Using Realized Volatility and ARCH type Models. Journal of Empirical Finance 11, 379-398.

Glosten, L.R., Jagannathan, R. and Runkle, D. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. Journal of Finance 48, 1779-1801.

Gomez, H.W., Torres, F.J. and H. Bolfarine. (2007). Large-sample inference for the epsilon skew-t distribution. Communications in Statistics-Theory and Methods, 36, 73-81.

Harvey, A.C. (2012). Exponential conditional volatility models. Unpublished manuscript (revised).

Harvey, A.C. and Chakravarty, T. (2008). Cambridge Working paper in Economics, CWPE 0840.

Harvey, C.R. and A. Siddique (2000). Conditional Skewness in Asset Pricing Tests, The Journal of Finance, 55, 1263-1295.

Jensen, S. T. and A. Rahbek (2004). Asymptotic inference for nonstationary GARCH. *Econometric Theory* **20**, 1203-26.

Laurent, S. (2009). G@RCH6. Timberlake Consultants Ltd., London.

Linton, O. (2008). ARCH models. In The New Palgrave Dictionary of Economics, 2nd Ed.

Ljung, G. and G. Box (1979). On a Measure of Lack of Fit in Time Series Models. Biometrika 66, 265–270.

McLeod, A.I., and W.K. Li (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* 4: 269-73.

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347-370

Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics 6, 461–464.

Slater, L.J. (1965). Confluent hypergeometric functions. In Abramowitz, M. and I. A. Stegun (Eds.), *Handbook of Mathematical Functions*, 503-35, Dover Publications Inc., New York.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *Annals of Statistics* **34**, 2449-2495.

Sucarrat, G. (2012). betategarch: Estimation and simulation of the first-order Beta-t-EGARCH model. R package version 1.2. `http://cran.r-project.org/web/packages/betategarch/`

Taylor, J. and A. Verbyla (2004). Joint modelling of location and scale parameters of the t distribution. *Statistical Modelling* 4, 91-112.

Zhang, X., Creal, D., Koopman, S.J. and A. Lucas (2011). Modeling Dynamic Volatilities and Correlations under Skewness and Fat Tails. Tinbergen Institute Discussion Paper 11-078/2.

Zhu, D. and J.W. Galbraith (2010). A generalized asymmetric Student-t distribution with application to financial econometrics. Journal of Econometrics 157, 297-305.

Zhu., D. and V. Zinde-Walsh (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, 148, 86-99.

Zivot, E. (2009). Practical issues in the analysis of univariate GARCH models. In Anderson, T.G. et al. Handbook of Financial Time Series, 113-155. Berlin: Springer-Verlag.

Appendix: Asymptotic properties of the ML estimator

This appendix explains how to derive the information matrix of the ML estimator for the first-order model and outlines a proof for consistency and asymptotic normality; full details can be found in Harvey (2012).

As noted in the text, if the model is to be identified, κ must not be zero or such that the constraint $b < 1$ is violated. A more formal statement is that the parameters should be interior points of the compact parameter space which will be taken to be $|\phi| < 1$, $|\omega| < \infty$ and $0 < \kappa < \kappa_u$, $\kappa_L < \kappa < 0$ where κ_u and κ_L are values determined by the condition $b < 1$.

The first step is to decompose the derivatives of the log density wrt $\boldsymbol{\psi}$ into derivatives wrt $\lambda_{t|t-1}$ and derivatives of $\lambda_{t|t-1}$ wrt $\boldsymbol{\psi}$, that is

$$\frac{\partial \ln f_t}{\partial \boldsymbol{\psi}} = \frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}}, \quad i = 1, 2, 3.$$

Since the scores $\partial \ln f_t / \partial \lambda_{t|t-1}$ are $IID(0, \sigma_u^2)$ and so do not depend on $\lambda_{t|t-1}$,

$$\begin{aligned} E_{t-1}\left[\left(\frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}}\right)\left(\frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}}\right)'\right] &= \left[E\left(\frac{\partial \ln f_t}{\partial \mu}\right)^2\right] \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}'} \\ &= \sigma_u^2 \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}'}. \end{aligned}$$

Thus the unconditional expectation requires evaluating the last term. In order to do this, the following definitions, which specialize to the expressions in (45), are needed:

$$\begin{aligned} a &= \phi + \kappa E\left(\frac{\partial u_t}{\partial \lambda}\right) \\ b &= \phi^2 + 2\phi\kappa E\left(\frac{\partial u_t}{\partial \lambda}\right) + \kappa^2 E\left(\frac{\partial u_t}{\partial \lambda}\right)^2 \geq 0 \\ c &= \kappa E\left(u_t \frac{\partial u_t}{\partial \lambda}\right) \end{aligned} \tag{45}$$

We also note that the first derivative of the conditional score is

$$\frac{\partial u_t}{\partial \lambda_{t|t-1}} = \frac{-2(\nu+1)(y_t-\mu)^2 \nu \exp(2\lambda_{t|t-1})}{(\nu \exp(2\lambda_{t|t-1}) + y_t - \mu)^2)^2} = -2(\nu+1) b_t (1 - b_t),$$

and since, like u_t, this depends only on a beta variable, it is also IID. Hence the distribution of u_t and its first derivative are independent of $\lambda_{t|t-1}$. All moments of u_t and $\partial u_t / \partial \lambda$ exist for the t-distribution and the expressions for a, b and c are as in (45).

The derivative of $\lambda_{t|t-1}$ wrt κ is

$$\frac{\partial \lambda_{t|t-1}}{\partial \kappa} = \phi \frac{\partial \mu_{t-1|t-2}}{\partial \kappa} + \kappa \frac{\partial u_{t-1}}{\partial \kappa} + u_{t-1}, \qquad t = 2, ..., T.$$

However,

$$\frac{\partial u_t}{\partial \kappa} = \frac{\partial u_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \kappa},$$

Therefore

$$\frac{\partial \lambda_{t|t-1}}{\partial \kappa} = x_{t-1}\frac{\partial \mu_{t-1|t-2}}{\partial \kappa} + u_{t-1} \tag{46}$$

where

$$x_t = \phi + \kappa \frac{\partial u_t}{\partial \lambda_{t|t-1}}, \quad t = 1, ..., T. \tag{47}$$

Taking conditional expectations of x_t gives

$$E_{t-1}(x_t) = \phi + \kappa E_{t-1}\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right) = \phi + \kappa E\left(\frac{\partial u_t}{\partial \mu}\right),$$

where the last equality follows because $\partial u_t/\partial \lambda_{t|t-1}$ is IID and so unconditional expectations can replace conditional ones. The unconditional expression defines the general expression for the quantity 'a' in (45).

When the process for $\lambda_{t|t-1}$ starts in the infinite past and $|a| < 1$, taking conditional expectations of the derivatives at time $t-2$, followed by unconditional expectations gives

$$E\left(\frac{\partial \lambda_{t|t-1}}{\partial \kappa}\right) = E\left(\frac{\partial \lambda_{t|t-1}}{\partial \phi}\right) = 0 \quad \text{and} \quad E\left(\frac{\partial \lambda_{t|t-1}}{\partial \omega}\right) = \frac{1-\phi}{1-a}.$$

The derivatives wrt ϕ and ω are found in a similar way.

To derive the information matrix, square both sides of (46) and take conditional expectations to give

$$\begin{aligned} E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \kappa}\right)^2 &= E_{t-2}\left(x_{t-1}\frac{\partial \mu_{t-1|t-2}}{\partial \kappa} + u_{t-1}\right)^2 \\ &= b\left(\frac{\partial \mu_{t-1|t-2}}{\partial \kappa}\right)^2 + 2c\frac{\partial \mu_{t-1|t-2}}{\partial \kappa} + \sigma_u^2. \end{aligned} \tag{48}$$

where b and c are as defined in (13). Taking unconditional expectations gives

$$E\left(\frac{\partial \lambda_{t|t-1}}{\partial \kappa}\right)^2 = bE\left(\frac{\partial \mu_{t-1|t-2}}{\partial \kappa}\right)^2 + 2cE\left(\frac{\partial \mu_{t-1|t-2}}{\partial \kappa}\right) + \sigma_u^2$$

and so, provided that $b<1$,

$$E\left(\frac{\partial \lambda_{t|t-1}}{\partial \kappa}\right)^2=\frac{\sigma_u^2}{1-b}.$$

Expressions for other elements in the information matrix may be similarly derived; see Harvey (2012). Fulfillment of the condition $b<1$ implies $|a|<1$. That this is the case follows directly from the Cauchy-Schwartz inequality $E(x_t^2)\geq [E(x_t)]^2$.

Consistency and asymptotic normality can be proved by showing that the conditions for Lemma 1 in Jensen and Rahbek (2004, p 1206) hold. The main point to note is that the first three derivatives of $\lambda_{t|t-1}$ wrt κ, ϕ and ω are stochastic recurrence equations (SREs); see Brandt (1986) and Straumann and Mikosch (2006, p 2450-1). The condition $b<1$ is sufficient[7] to ensure that they are strictly stationarity and ergodic at the true parameter value. Similarly $b<1$ is sufficient to ensure that the squares of the first derivatives are strictly stationary and ergodic.

Let $\boldsymbol{\psi}_0$ denote the true value of $\boldsymbol{\psi}$. Since the score and its derivatives wrt μ in the static model possess the required moments, it is straightforward to show that (i) as $T\rightarrow\infty$, $(1/\sqrt{T})\partial \ln L(\boldsymbol{\psi}_0)/\partial \boldsymbol{\psi} \rightarrow N(0,\mathbf{I}(\boldsymbol{\psi}_0))$, where $\mathbf{I}(\boldsymbol{\psi}_0)$ is p.d. and (ii) as $T\rightarrow\infty$, $(-1/T)\partial^2 \ln L(\boldsymbol{\psi}_0)/\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}' \xrightarrow{P} \mathbf{I}(\boldsymbol{\psi}_0)$. The final condition in Jensen and Rahbek (2004) is concerned with boundedness of the third derivative of the log-likelihood function in the neighbourhood of $\boldsymbol{\psi}_0$. The derivatives of u_t, as well as u_t itself, are affine functions of terms of the form $b_t^*=b_t^h(1-b_t)^k$, where h and k are non-negative integers. Since

$$b_t=h(y_t;\boldsymbol{\psi})/(1+h(y_t;\boldsymbol{\psi})),\qquad 0\leq h(y_t;\boldsymbol{\psi})\leq\infty,$$

where $h(y_t;\boldsymbol{\psi})$ depends on y_t and $\boldsymbol{\psi}$, it is clear that, for any admissible $\boldsymbol{\psi}$, $0\leq b_t\leq 1$ and so $0\leq b_t^*\leq 1$. Furthermore the derivatives of $\lambda_{t|t-1}$ must be bounded at $\boldsymbol{\psi}_0$ since they are stable SREs which are ultimately dependent on u_t and its derivatives. They must also be bounded in the neighbourhood of $\boldsymbol{\psi}_0$ since the condition $b<1$ is more than enough to guarantee the stability condition $E(\ln|x_t|)<0$.

Unknown shape parameters, including degrees of freedom, pose no prob-

[7] The necessary condition for strict stationarity is $E(\ln|x_t|)<0$. This condition is satisfied at the true parameter value when $|a|<1$ since, from Jensen's inequality, $E(\ln|x_t|)\leq \ln E(|x_t|)<0$ and as already noted $b<1$ implies $|a|<1$.

lem as the third derivatives (including cross-derivatives) associated with them are almost invariably non-stochastic.